DOCUMENT RESUME

ED 181 061 TM 009 969

AUTHOR Sympson, James B.
TITLE The Assessment of Basic Competencies: A New Test Battery.
PUB DATE 3 Sep 79
NOTE 29p.: Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes: Cognitive Tests: Elementary Education: Equated Scores: Intellectual Development: Item Analysis: Language Skills: *Logical Thinking: Mathematical Concepts: Mathematical Models: *Norms: Preschool Education: Probability: Sampling: *Scoring: Scoring Formulas: Simulation: *Student Ability: Technical Reports: *Test Construction: Test Results
IDENTIFIERS *Assessment of Basic Competencies: Interval Scaling: *Latent Trait Model: Three Parameter Model

ABSTRACT

Development of The Assessment of Basic Competencies (ABC), a test battery based on the three-parameter logistic model, is described. Eleven dimensions of intellectual growth are measured, from the pre-kindergarten level through ninth grade. An educationally relevant skill domain is represented by each test. Unique properties of the test, based on item characteristic curve theory, are reviewed and two procedures utilized in developing the battery are described. The methods used to place each test's items on a common metric and to generate scoring tables for each test are described. Overlapping subtests or overlapping groups of testees are not required, and hand-calculated test scores can be converted into a maximum likelihood point estimate and a distribution-free interval estimate of the testee's ability. (Author/MH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED181061

# The Assessment of Basic Competencies
## A New Test Battery

James B. Sympson
Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James B. Sympson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

APA 1979 Annual Meeting
New York City
September 3, 1979

TM009 969

## Abstract

Development of a new test battery based on the three-parameter logistic model is described. The battery measures eleven dimensions of intellectual growth from the pre-kindergarten level through ninth grade. Each test's content was selected to represent an educationally relevant skill domain. The battery allows an educator or school psychologist to estimate a child's location on eleven equal-interval developmental scales and also provides information useful for the design of individualized instructional programs. The tests are individually administered, and test items are selected adaptively in order to tailor the test to each child.

The discussion does not emphasize the content of the battery. Rather, several unique properties of tests based on item characteristic curve theory are reviewed and two novel procedures utilized in developing the battery are described. These properties and procedures can be useful in a variety of measurement applications. In particular, the possibility of measuring change over time on an equal-interval scale, the possibility of generating comparable test scores for people who have taken different sets of items, and the possibility of obtaining both norm-referenced and criterion-referenced interpretations of test performance are discussed.

With regard to procedures used in developing the battery, the method used to place each test's items on a common metric and the method used to generate scoring tables for each test are described. The procedure

for placing items on a common metric is unusual in that it does not require either overlapping subtests or overlapping groups of testees. The scoring tables for the battery are unusual in that they allow the test administrator to convert a simple hand-calculated test score into a maximum likelihood point estimate and a distribution-free interval estimate of the testee's ability. This has not previously been possible with three-parameter item response models.

# The Assessment of Basic Competencies: A New Test Battery

James B. Sympson

Educational Testing Service

Today I will discuss some features of a new test battery that was developed using the three-parameter logistic latent trait model (Birnbaum, 1968, p. 405). Development of this test battery, the Assessment of Basic Competencies (ABC), was conceived and funded by Dr. J. P. Somwaru of the Minnesota State Department of Education. I served as psychometric and statistical consultant for this privately funded project while affiliated with the University of Minnesota in Minneapolis. The ABC battery is scheduled for publication by Scholastic Testing Service, Inc., of Bensenville, Illinois, later this year.

The ABC measures eleven dimensions of intellectual growth from the pre-kindergarten level through ninth grade. Each scale's content was selected by Dr. Somwaru to represent an educationally relevant skill domain. The skill domains are logically grouped into three broad clusters: Information Processing Skills, Language Skills, and Mathematical Reasoning Skills The three scales associated with the Information Processing cluster are labeled Relating, Organizing, and Observing. The five scales associated with the Language cluster are labeled Reading, Decoding, Comprehending Expressions, Producing Expressions, and Understanding Words. The three scales associated with the Mathematical Reasoning cluster are Solving Problems, Understanding

Concepts, and Number and Operations. The skills assessed in each ABC scale are briefly summarized in Figure 1.

---

Insert Figure 1 about here

---

Each test in the ABC is individually administered in a manner similar to the Stanford-Binet (Anastasi, 1968, chap. 8). The items in each scale are ordered in terms of their latent trait difficulty parameters and a starting item is selected by the test administrator on the basis of the child's current grade-level assignment. Assuming the first item is passed, items of increasing difficulty are administered until either five items in a row are missed or the most difficult item in the scale is administered. Then, if five items in a row were not passed at some previous point, items of decreasing difficulty below the starting item are administered until the basal item (the most difficult of five items passed in a row) is identified. At this point, testing for the scale is terminated and testing with another scale starts. When testing is completed, a hand-calculated test score is computed for each scale. Each of these scores is then converted into a latent trait ability estimate and lower and upper 75% confidence limits for the ability estimate.

Latent trait ability estimates are estimates of a person's position on an inferred interval-level ability scale. These estimates have the advantage of being essentially independent of the particular items administered. Thus, a person can be administered two sets of items

at two points in time, and the difference between the resulting ability estimates will be a direct interval-level measure of cognitive growth. Moreover, the ability estimates of two people who have taken different sets of items are directly comparable.

Since a person's estimated ability score can be used in conjunction with the latent trait parameters of one or more test items to calculate the probability of a correct response to each of these items, substantive statements can be made about a person's expected level of performance (e.g., their expected proportion correct) in a set of items that they have not taken previously. This provides the type of interpretation sought in content-referenced and criterion-referenced testing. Also, since latent trait ability estimates can be converted into standard scores or percentile ranks relative to any norm group for which a distribution of estimated ability scores has been obtained, the advantages of norm-referenced score interpretations are available.

---

Insert Figure 2 about here

---

The ABC score report form shown in Figure 2 provides space for recording and graphing both latent trait ability estimates and various norm-referenced scores. The developmental age score for each scale is equal to the mean age in months of persons in the calibration and norming sample who obtained a given ability estimate. This is not the same as a mental age score. The latter is an age equivalent score (Anastasi, 1968, p.46).

Discussion of the rationale behind the content selected for the ABC and potential applications of the test battery can be found in papers that have been prepared by Dr. Somwaru (1979a, 1979b).

Design for Data Collection

Seven of the scales in the ABC contain 100 items that range in difficulty from a level suitable for children at age 3+ years to a level suitable for ninth grade students. Four of the scales contain 85 items ranging in difficulty from the kindergarten level to the ninth grade level. The items included in an ABC scale were selected from among either 200 or 180 items that had been written for the scale and calibrated with the latent trait item parameterization program LOGIST (Wingersky and Lord, 1973).

---

Insert Table 1 about here

---

The data collection plan for a single ABC scale is shown in Table 1. The set of 200 or 180 items prepared for an ABC scale were initially rank ordered by Dr. Somwaru in terms of their perceived subject matter difficulty level. Then, twenty-item subtests ranging from low to high difficulty were formed and each subtest (A through J or B through J) was administered to 60-70 students at the grade level for which the subtest's difficulty seemed most appropriate. In addition, each subtest was administered to 60-70 students from the grade levels immediately below and above the most appropriate grade.

Thus, a total of 180-210 individuals responded to the items in each scale subtest. Each individual in the calibration and norming sample was administered one subtest from each of two different scales. A total of 2120 items were calibrated using the responses of over 10,000 individuals.

The responses of the individuals who took each subtest were used by LOGIST to parameterize the items included in the subtest. The lower asymptote parameter was set to zero for free response items and to 1/(2A), where A is the number of response alternatives, for multiple choice items. Within each subtest the item discrimination parameters were assumed to be equal and the item difficulty parameters were estimated independently for each item. It is important to note that no assumption was made that the item discrimination parameters were constant over the entire ability range from pre-kindergarten to ninth grade levels. We only assumed that items in the relatively narrow difficulty range spanned by each subtest were equally discriminating. The general level of item discrimination was left free to increase, decrease, or remain constant over the range of cognitive development covered by each scale.

Procedure for Linking Item Parameters

The latent trait item parameterization procedure that was used to generate item parameters for the ABC expressed these parameters relative to the ability distribution of the particular group that completed each subtest. Since the groups that completed the subtests associated with a given scale differed with regard to their level and

range of ability, the item parameters for the items developed for each scale were not initially expressed in terms of a common metric. Thus, a single metric that spanned the entire ability range had to be created. The following procedure accomplished this goal.

First, we assumed that each group of 60-70 students was a random sample from a particular grade-level population. Thus, using fourth graders as an example, the 3 groups of fourth grade students that took scale subtests E, F, and G were assumed to be 3 random samples from the same population. This same type of assumption was made for all 12 grade levels.

Given this assumption, we may note that the 120-140 person group comprised of third and fourth graders that took subtest E of a scale must be a random sample from the same "composite" population as the 120-140 third and fourth graders that took subtest F. Similarly, the 120-140 fourth and fifth graders that took subtest F of a scale and the 120-140 fourth and fifth graders that took subtest G must be two random samples from the same composite population. Hence, except for sampling error, each such pair of 120-140 person samples from a common composite population must have the same underlying level and range of ability. We used this fact to achieve a "linking" of the item parameters.

The people who took scale subtest F were selected as an initial reference group and then all 200 (or 180) items' parameters were expressed relative to this group as follows. First, we noted the mean ability level among third and fourth graders who took subtest F. This quantity may be designated $M[{}_F\hat{\theta}_{3,4}]$. Here, the symbol $\hat{\theta}$ stands for the estimated ability of an individual, the leading subscript indicates

that we are referring to people who took subtest F, and the trailing subscripts indicate that we refer only to third and fourth graders. Next, we calculated $M[{}_E\hat{\theta}_{3,4}]$ and the two standard deviations $S[{}_F\hat{\theta}_{3,4}]$ and $S[{}_E\hat{\theta}_{3,4}]$. We then brought the 180-210 students who took subtest E onto the metric of the subtest F reference group using the linear transformation shown in Equation 1.

$$ {}_E\hat{\theta}'_{2,3,4} = \left(\frac{S[{}_F\hat{\theta}_{3,4}]}{S[{}_E\hat{\theta}_{3,4}]}\right) ({}_E\hat{\theta}_{2,3,4} - M[{}_E\hat{\theta}_{3,4}]) + M[{}_F\hat{\theta}_{3,4}] \qquad [1] $$

The initial item difficulty parameters (${}_Eb$) and discrimination parameters (${}_Ea$) for the items in subtest E were then expressed relative to the subtest F metric with the transformations shown in Equations 2 and 3.

$$ {}_Eb' = \left(\frac{S[{}_F\hat{\theta}_{3,4}]}{S[{}_E\hat{\theta}_{3,4}]}\right) ({}_Eb - M[{}_E\hat{\theta}_{3,4}]) + M[{}_F\hat{\theta}_{3,4}] \qquad [2] $$

$$ {}_Ea' = \left(\frac{S[{}_E\hat{\theta}_{3,4}]}{S[{}_F\hat{\theta}_{3,4}]}\right) {}_Ea \qquad [3] $$

In precisely the same manner, the 180-210 students who completed subtest G were brought onto the metric of the subtest F reference group using the transformation of Equation 4 and the subtest G item parameters

were expressed relative to the subtest F metric with the transformations shown in Equations 5 and 6.

$$G\hat{\theta}'_{4,5,6} = \left(\frac{S[F\hat{\theta}_{4,5}]}{S[G\hat{\theta}_{4,5}]}\right) \left(G\hat{\theta}_{4,5,6} - M[G\hat{\theta}_{4,5}]\right) + M[F\hat{\theta}_{4,5}] \qquad [4]$$

$$G^{b'} = \left(\frac{S[F\hat{\theta}_{4,5}]}{S[G\hat{\theta}_{4,5}]}\right) \left(G^{b} - M[G\hat{\theta}_{4,5}]\right) + M[F\hat{\theta}_{4,5}] \qquad [5]$$

$$G^{a'} = \left(\frac{S[G\hat{\theta}_{4,5}]}{S[F\hat{\theta}_{4,5}]}\right) G^{a} \qquad [6]$$

Next we brought the 180-210 people who completed subtest D onto the metric of the subtest F reference group using Equation 7.

$$D\hat{\theta}'_{1,2,3} = \left(\frac{S[E\hat{\theta}'_{2,3}]}{S[D\hat{\theta}_{2,3}]}\right) \left(D\hat{\theta}_{1,2,3} - M[D\hat{\theta}_{2,3}]\right) + M[E\hat{\theta}'_{2,3}] \qquad [7]$$

The initial parameter values for the items in subtest D were then expressed relative to the subtest F metric metric using Equations 8 and 9. (Note that Equations 7, 8, and 9 are functions of the transformed scores obtained with Equation 1.)

$$D^{b'} = \left(\frac{S[_E\hat{\theta}'_{2,3}]}{S[_D\hat{\theta}_{2,3}]}\right) \left(_Db - M[_D\hat{\theta}_{2,3}]\right) + M[_E\hat{\theta}'_{2,3}] \quad [8]$$

$$D^{a'} = \left(\frac{S[_D\hat{\theta}_{2,3}]}{S[_E\hat{\theta}'_{2,3}]}\right) D^a \quad [9]$$

Similarly, the ability scores of people who completed subtest H and the parameters of the items in that subtest were transformed with Equations 10,11, and 12.

$$H\hat{\theta}'_{5,6,7} = \left(\frac{S[_G\hat{\theta}'_{5,6}]}{S[_H\hat{\theta}_{5,6}]}\right) \left(_H\hat{\theta}_{5,6,7} - M[_H\hat{\theta}_{5,6}]\right) + M[_G\hat{\theta}'_{5,6}] \quad [10]$$

$$H^{b'} = \left(\frac{S[_G\hat{\theta}'_{5,6}]}{S[_H\hat{\theta}_{5,6}]}\right) \left(_Hb - M[_H\hat{\theta}_{5,6}]\right) + M[_G\hat{\theta}'_{5,6}] \quad [11]$$

$$H^{a'} = \left(\frac{S[_H\hat{\theta}_{5,6}]}{S[_G\hat{\theta}'_{5,6}]}\right) H^a \quad [12]$$

This procedure was applied recursively, proceeding from the middle of the grade range to the two tails, until all the ability estimates and all the item parameters for a scale were expressed relative to the metric of the subtest F reference group. At this point, except for small inaccuracies due to sampling error, a continuous, interval-level scale of ability had been created that spanned the range from the 3+ (or 4+) level to the 9th grade level. (The claim that inaccuracies due to sampling errors will be tolerably small with 70 cases per cell was supported by a computer simulation of this procedure that was conducted by the author.)

While the goal of creating a common metric for all items was accomplished by the foregoing procedure, a further scale transformation was executed. Since we wished to express all item difficulties and person abilities as positive integers between 100 and 200, the ability scale was transformed so that the mean ability estimate in the kindergarten group was 140 and the mean ability estimate in the 8th grade group was 190. This was accomplished by applying the transformations shown in Equations 13, 14, and 15 to the calibration sample ability estimates and item parameters.

$$\hat{\theta}'' = \left( \frac{50}{M[\hat{\theta}'_8] - M[\hat{\theta}'_K]} \right) (\hat{\theta}' - M[\hat{\theta}'_K]) + 140 \qquad [13]$$

$$b'' = \left( \frac{50}{M[\hat{\theta}'_8] - M[\hat{\theta}'_K]} \right) (\hat{b}' - M[\hat{\theta}'_K]) + 140 \qquad [14]$$

$$a'' = \left( \frac{M[\hat{\theta}_8'] - M[\hat{\theta}_K']}{50} \right) a' \qquad [15]$$

In these equations, $M[\hat{\theta}_K']$ is the mean ability estimate in the kindergarten group (prior to the final transformation) and $M[\hat{\theta}_8']$ is the mean ability estimate in the 8th grade group.

The procedures described above were executed for each of the scales in the ABC test battery. The result was 11 interval-level measurement scales that can be used to assess growth over time on 11 fundamental dimensions of intellectual competency.

Generation of Scoring Tables

Ability estimation in a hand-scored test based on the three-parameter logistic latent trait model has previously been impossible. This was due to lack of a suitable procedure for generating scoring tables for such tests. While I was still at the University of Minnesota, I developed a very general procedure for generating scoring tables (Sympson, 1979) and have used it in constructing the scoring tables for the eleven scales in the ABC battery. This procedure is summarized below.

Latent trait models are stochastic models that express the probability of a correct response to a test item as a function of the item's parameters and the ability level of the person tested. Once the latent trait model, the item's parameters, and a particular level of ability are specified, the conditional sampling distribution of the

item response (correct or incorrect) can be generated. This is a binomial distribution. In a similar manner, for a given model and a particular set of test items, there is a conditional sampling distribution of the number correct score at each possible ability level. This distribution is a compound-binomial distribution (Lord & Novick, 1968, p. 384).

More generally, for any particular testing and scoring procedure, there is a conditional distribution of test scores associated with each possible latent trait ability level. For some testing and scoring procedures, these conditional distributions can be generated, or closely approximated, using mathematically derived functions. In the case of adaptive tests, however, where different people receive different items and possibly even different numbers of items, the mathematical approach to generating or approximating conditional test score distributions is intractable.

When analytical methods cannot be used to derive conditional test score distributions, an alternative approach is to program a computer to administer a large number of simulated tests at a fixed ability level, and then use the obtained distribution of test scores as an estimate of the unknown conditional sampling distribution. This type of simulation program incorporates a response generator that "responds" to items selected by the testing procedure. The responses are ones and zeroes, indicating correct and incorrect answers respectively. In each case, the probability that a correct answer will be generated is determined by inserting the item's parameters

and the latent trait ability level into the equation for the item characteristic curve.

Given that a procedure is available for generating estimated sampling distributions of test scores, one can produce a large number of such distributions at closely spaced ability levels. The ability levels selected should span the range of ability that will be encountered during testing. These distributions can then be arrayed in a matrix with each row corresponding to a particular test score (or short interval of test scores) and each column corresponding to a particular level of ability at which a conditional distribution of test scores was generated. A schematic representation of such a matrix is shown in Table 2.

---

Insert Table 2 about here

---

The elements of the matrix shown in Table 2 are conditional probabilities of the type $P(X_i|\theta_j)$ where $X_i$ represents a particular score, or a short score interval, and $\theta_j$ represents a particular ability level. If one examines the $P(X_i|\theta_j)$ values in a given row of the matrix and identifies the highest of these values, the ability level associated with this highest conditional probability is approximately equal to the maximum likelihood estimate of ability for people whose test scores fall in the score interval identified by $X_i$. Of course, since the matrix does not contain an infinite number of columns, the ability

level corresponding to the maximum $P(X_i|\theta_j)$ value in a row is only a rough approximation to the exact maximum likelihood point estimate. However, if the ability levels at which the conditional distributions have been generated are close together, a very good approximation to the exact point estimate can be obtained by passing a quadratic function through the $P(X_i|\theta_j)$ values near the row maximum and determining the ability level that corresponds to the highest point of this function. This amounts to quadratic interpolation between the columns of the matrix of conditional probabilities.

It is important to note that the statistical efficiency of the maximum likelihood estimates obtained via the foregoing procedure will depend on the nature of the test score, X. If X is a sufficient statistic (Birnbaum, 1968, p. 426) for estimating ability under the particular latent trait model involved, then the maximum likelihood point estimate derived from X will be asymptotically efficient (Birnbaum, 1968, p.455). If X is not a sufficient statistic, then the maximum likelihood point estimate will not be as efficient an estimator as it would be if derived from either the original vector of item responses or some sufficient statistic obtained from that response vector.

In general, there is no test score that can be calculated by hand which is a sufficient statistic for estimating ability under a three-parameter logistic latent trait model (Birnbaum, 1968, p. 459). Thus, hand scoring of a test that contains multiple-choice items will necessarily reduce the efficiency of ability estimates relative to the efficiency obtainable with computer scoring of the test. However, the

procedure described above provides a method for minimizing this loss of efficiency due to hand scoring. This procedure extracts all the information about a person's ability level that a given hand-calculated test score contains.

In addition to point estimates, this procedure also provides interval estimates of ability. These are obtained as follows: Within each column of the matrix of conditional test score distributions, the set of smallest $P(X_i|\theta_j)$ values whose sum is equal to or just less than $\alpha$ (where $0 < \alpha < 1$) is called the $\alpha$-level "rejection" set. The remaining elements in each column are called the $(1-\alpha)$-level "acceptance" set. All the elements in the matrix that fall within an acceptance set are flagged with a distinctive marker. Now, when one reads along any specified row of the matrix, some of the elements in the row will be flagged and some will not.

The set of ability levels that have flagged elements in a specified row of the matrix are points that fall within the $100(1-\alpha)\%$ confidence interval (Sympson, 1979; Kendall & Stuart, 1973, chap. 20) for estimating the ability of people whose test score is associated with that row. We assume that ability levels between the observed points are also in the confidence interval and that the interval limits fall approximately half-way between the most extreme flagged elements and the nearest non-flagged elements. Confidence intervals constructed in this manner will capture a person's true ability level between the interval limits at least $100(1-\alpha)\%$ of the time if the person is tested repeatedly using parallel item pools (i.e., one-to-one matching of parameters in the series of pools).

The confidence intervals obtained from this procedure are not equivalent to so-called "confidence intervals" calculated using the conditional standard error of $\hat{\theta}$ given $\theta$ (i.e., the conditional standard error of measurement). In latent trait theory, the conditional standard error of $\hat{\theta}$ is seen to vary over ability levels as a function of the characteristics of the testing and scoring procedure. However, it can be shown (Sympson, 1979) that the procedure which uses the standard error of $\hat{\theta}$ to construct confidence intervals assumes that conditional distributions of $\hat{\theta}$ are normal with constant variance over all ability levels. Such an assumption is generally unrealistic and, in light of the availability of the procedure described here, unnecessary.

The fact that this procedure provides interval estimates of ability has another benefit beyond the obvious one of supplementing the maximum likelihood point estimate. The procedure can be used with a given test scored in various ways in order to determine which scoring method results in the shortest confidence intervals for ability estimates. A scoring method that provides uniformly shorter confidence intervals is obviously preferable.

In the case of the ABC test battery, the initial plan was to score each test by calculating the average difficulty of items answered correctly. Examination of the confidence intervals that resulted with this test score showed that it had some undesirable characteristics, particularly at low ability levels when guessing was possible. As a result of this finding, other scoring methods were considered and the most promising of these, the "pseudo" number correct score, was tried.

This score is calculated by adding the number of items answered correctly above the basal item to the item number of the basal item. This score was found to provide shorter confidence intervals than the average difficulty score at all ability levels. Thus, in each ABC test it is the pseudo number correct score that is converted to a maximum likelihood point estimate and confidence limits for $\theta$.

One problem that arose in the development of scoring tables for the ABC test battery was the large amount of time it would take to generate a different scoring table for every possible test entry point for each test in the battery. Preliminary analyses indicated that it would be sufficient to run simulations at several entry points for each test and then interpolate linearly between the obtained scoring tables to obtain tables for the other possible entry points. Therefore, at test entry points corresponding to items 1, 11, 21, ..., 91 in the 100 item scales and items 1, 11, 21, ..., 81 in the 85 item scales, the foregoing procedure was used to obtain the scoring table (maximum likelihood point estimates and confidence limits) for each entry point. Treating the point estimates, the lower confidence limits, and the upper confidence limits separately, an array with either ten or nine columns, corresponding to the above entry points, and either 101 or 86 rows, corresponding to possible pseudo number correct scores, was formed. This array of point estimates or confidence limits was "smoothed" by computer to eliminate irregularities due to sampling error and then was expanded to either a 91 column array or an 81 column array by linear interpolation between the elements of the original columns.

For each test in the battery, the result of the foregoing procedure was a scoring table that provides maximum likelihood point estimates and lower and upper 75% confidence limits for estimating the ability of people who start a test at a particular item number and who obtain a particular pseudo number correct score. In generating these scoring tables, the statistical equivalent of over 5,000 test administrations was used to estimate the sampling distribution of test scores for individuals with a specified level of ability who start a test at a given entry point. Since sampling distributions were generated at ability levels ranging from 50 points below each entry-point difficulty value to 50 points above each entry-point difficulty value, the statistical equivalent of over 500,000 test administrations was used to obtain the point and interval estimates associated with a single test entry point. Finally, since each 100 item scale's scoring table was derived from the results obtained at ten test entry points, we see that each of the scoring tables for the 100 item scales in the ABC is based on the statistical equivalent of more than 5 million test administrations. Similarly, the scoring table for each of the 85 item scales in the ABC test battery is based on the statistical equivalent of more than 4,500,000 test administrations.

## References

Anastasi, A. Psychological testing (3rd ed.). New York: Macmillan, 1968.

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968, Chapters 17-20.

Kendall, M. G., & Stuart, A. The advanced theory of statistics (Vol. 2, 3rd ed.). London: Griffin, 1973.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). Minneapolis: University of Minnesota, Department of Psychology, Institute for Research on Learning Disabilities, April 1979 (a).

Somwaru, J. P. Assessment of Basic Competencies: An alternative to intelligence tests. Unpublished manuscript, 1979 (b). (Available from the author at: Division of Special Services, Minnesota Department of Education, 550 Cedar St., St. Paul, MN 55101.)

Sympson, J. B. An approach to some estimation problems in latent trait theory (Research Report 79-X). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program (in preparation).

Wingersky, M. S., & Lord, F. M. A computer program for estimating examinee ability and item characteristic curve parameters when there are omitted responses (Research Memorandum 73-2). Princeton, N.J.: Educational Testing Service, 1973.

| Subtests | Items Included | Age | | Grade | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 3+ | 4+ | K | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | 1-20 | X | X | X | | | | | | | | | |
| B | 21-40 | | X | X | X | | | | | | | | |
| C | 41-60 | | | X | X | X | | | | | | | |
| D | 61-80 | | | | X | X | X | | | | | | |
| E | 81-100 | | | | | X | X | X | | | | | |
| F | 101-120 | | | | | | X | X | X | | | | |
| G | 121-140 | | | | | | | X | X | X | | | |
| H | 141-160 | | | | | | | | X | X | X | | |
| I | 161-180 | | | | | | | | | X | X | X | |
| J | 181-200 | | | | | | | | | | X | X | X |

Note: Each X represents 60-70 students drawn at random.
Four ABC scales had no subtest A.

Table 1

Sampling Plan for a Single ABC Scale

| | | | | | |
|---|---|---|---|---|---|
| $P(X_{15}\|\theta_1)$ | $P(X_{15}\|\theta_2)$ | | $P(X_{15}\|\theta_{51})$ | $P(X_{15}\|\theta_{100})$ | $P(X_{15}\|\theta_{101})$ |
| $P(X_{14}\|\theta_1)$ | $P(X_{14}\|\theta_2)$ | | $P(X_{14}\|\theta_{51})$ | $P(X_{14}\|\theta_{100})$ | $P(X_{14}\|\theta_{101})$ |
| $P(X_{13}\|\theta_1)$ | $P(X_{13}\|\theta_2)$ | | $P(X_{13}\|\theta_{51})$ | $P(X_{13}\|\theta_{100})$ | $P(X_{13}\|\theta_{101})$ |
| $P(X_{12}\|\theta_1)$ | $P(X_{12}\|\theta_2)$ | | $P(X_{12}\|\theta_{51})$ | $P(X_{12}\|\theta_{100})$ | $P(X_{12}\|\theta_{101})$ |
| $P(X_{11}\|\theta_1)$ | $P(X_{11}\|\theta_2)$ | | $P(X_{11}\|\theta_{51})$ | $P(X_{11}\|\theta_{100})$ | $P(X_{11}\|\theta_{101})$ |
| $P(X_{10}\|\theta_1)$ | $P(X_{10}\|\theta_2)$ | | $P(X_{10}\|\theta_{51})$ | $P(X_{10}\|\theta_{100})$ | $P(X_{10}\|\theta_{101})$ |
| $P(X_9\|\theta_1)$ | $P(X_9\|\theta_2)$ | | $P(X_9\|\theta_{51})$ | $P(X_9\|\theta_{100})$ | $P(X_9\|\theta_{101})$ |
| $P(X_8\|\theta_1)$ | $P(X_8\|\theta_2)$ | . . . | $P(X_8\|\theta_{51})$ | $P(X_8\|\theta_{100})$ | $P(X_8\|\theta_{101})$ |
| $P(X_7\|\theta_1)$ | $P(X_7\|\theta_2)$ | . . . | $P(X_7\|\theta_{51})$ | $P(X_7\|\theta_{100})$ | $P(X_7\|\theta_{101})$ |
| $P(X_6\|\theta_1)$ | $P(X_6\|\theta_2)$ | | $P(X_6\|\theta_{51})$ | $P(X_6\|\theta_{100})$ | $P(X_6\|\theta_{101})$ |
| $P(X_5\|\theta_1)$ | $P(X_5\|\theta_2)$ | | $P(X_5\|\theta_{51})$ | $P(X_5\|\theta_{100})$ | $P(X_5\|\theta_{101})$ |
| $P(X_4\|\theta_1)$ | $P(X_4\|\theta_2)$ | | $P(X_4\|\theta_{51})$ | $P(X_4\|\theta_{100})$ | $P(X_4\|\theta_{101})$ |
| $P(X_3\|\theta_1)$ | $P(X_3\|\theta_2)$ | | $P(X_3\|\theta_{51})$ | $P(X_3\|\theta_{100})$ | $P(X_3\|\theta_{101})$ |
| $P(X_2\|\theta_1)$ | $P(X_2\|\theta_2)$ | | $P(X_2\|\theta_{51})$ | $P(X_2\|\theta_{100})$ | $P(X_2\|\theta_{101})$ |
| $P(X_1\|\theta_1)$ | $P(X_1\|\theta_2)$ | | $P(X_1\|\theta_{51})$ | $P(X_1\|\theta_{100})$ | $P(X_1\|\theta_{101})$ |
| $\theta_1$ | $\theta_2$ | | $\theta_{51}$ | $\theta_{100}$ | $\theta_{101}$ |

Table 2

Schematic Representation of Matrix of Conditional Probabilities $P(X_i|\theta_j)$

## DEFINITION OF BASIC COMPETENCIES

### LANGUAGE SKILLS

READING:
- Literal and inferential comprehension-detail, main idea, sequence, cause-effect.

DECODING:
- Knowing letters and sounds, discriminating between sounds, pronouncing words.

COMPREHENDING and PRODUCING EXPRESSIONS:
- Using nouns, adjectives, adverbs, pronouns.
- Using verbs, auxiliary, modal, copula.
- Using function words – prepositions, determiners, conjunctions, qualifiers, etc.
- Understanding sentence structure.
- Transforming sentences.
- Knowing usage and idioms.

UNDERSTANDING WORDS:
- Words for objects, descriptions, actions, categories, and abstract ideas.

### MATH. REASONING SKILLS

SOLVING PROBLEMS:
- Solving problems involving various applications in 1, 2, or 3 steps.

UNDERSTANDING CONCEPTS:
- Sets and relations.
- Number sentences and symbolic expressions.
- Units of measurement – English and metric.
- Units of money and combinations.
- Geometry and graphs.

KNOWING NUMBER AND OPERATIONS:
- Number and numeration.
- Order
- Four operations with whole numbers, common and decimal fractions.
- Conservation of number.

### INFORMATION PROCESSING SKILLS

RELATING:
- Identifying incongruent relations, inferring rules, completing analogies, solving combinations, making transitive inference, solving problems.

ORGANIZING:
- Excluding from class, including in class, relating classes, ordering in sequence.

OBSERVING:
- Comparing for similarity or difference, seeing and remembering, completing patterns, visualizing figures after rotation, integrating eye-hand activity, finding figure in background, conserving mass, length, weight, volume.

Figure 1

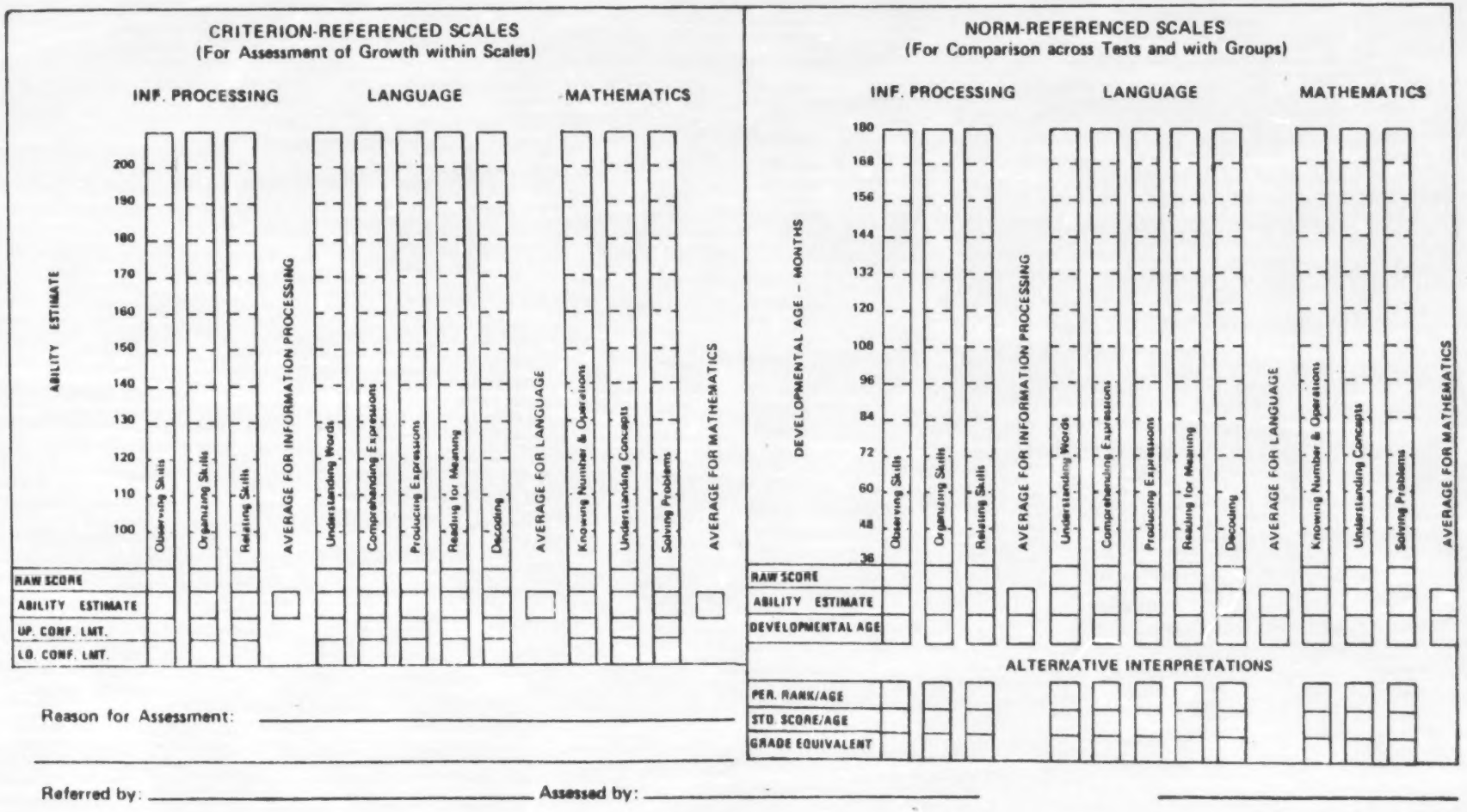
(Abstracted from User's Manual, Assessment of Basic Competencies. Copyright by J. P. Somwaru)

## ASSESSMENT OF BASIC COMPETENCIES

Name: _____ Date of Testing: Yr. ____ Mth. ____ Day ____

School: _____ Date of Birth: Yr. ____ Mth. ____ Day ____

Teacher: _____ Grade: ____ Age (Nearest Mth.) _____ Sex: ____




Reason for Assessment: _____

Referred by: _____ Assessed by: _____

Figure 2

(Abstracted from User's Manual, Assessment of Basic Competencies. Copyright by J. P. Sonwaru)